Informatics Institute of Technology In collaboration with University of Westminster, UK.

“Commit analyzing tool for Git version control system”

A Dissertation By Hasini Nimesha Wijerthna

Supervised by Mr. Saman Hettiarachchi

Submitted in partial fulfillment of the requirements for the BEng (Hons) Software Engineering Degree Department of Computing

May 2018

Abstract

Git is a distributed version control system allows each developer in a software project to have their local copy of the project that they have been working on. The transparent mechanism behind the version control system is to associate everyone using pull requests and commit messages which also can be reviewed later. This vast amount of activity could relate to implement research models can be derived for better enable productivity and communication amongst Git developers.

When developers are involved in large repositories with many edits throughout the project it is crucial to commit with a relevant commit message to ensure every member in the project to aware about the associated changes of the project. Since these commits make an informed overall decision about additional modification. The existing system allows developers to filter commits by the commit message and manually filter those code changes. If this procedure doing in projects with higher number of commits this may not be efficient and regardless the only way to identify additions and deletions of files, libraries and plugins is to manually checking the dependency installation files which not be efficient in large collection of files to check these changes manually. Most importantly this procedure of maintaining the code base varies from developer to developer.

In large repositories with many commits throughout the repository, Commit Analyzer was designed and implemented to be able to flag the commits and pull requests which are crucial to ensure everyone to be aware of the changes associated with a commit and can make an informed decision about whether to merge or ask for additional modifications in major milestones such as in a software release. So this model retrieves crucial details of source code changes and provides a summarized report in a given period of time.

System accuracy was thoroughly tested based on the rules while accuracy was tested based on the output values by Commit Analyzer. The usage of Paragraph vector classifier makes Commit Analyzer perform well. Commit Analyzer produces acceptable results and is hence justified. The tested results attested that the analysis, design, implementation and documentation have been carried out in an efficient manner.

Subject Descriptors: I.2.11 Distributed Artificial Intelligence, F.4.1 Mathematical Logic

Key words: Paragraph vector classifying, Word2vec two-layer neural net, Commit analyzing, Categorization